AUTOMATIC MEASUREMENT OF AUDIO PRESENCE AND LEVEL BY DIRECT
PROCESSING OF AN MPEG DATA STREAM

Inventors:
Thomas H. James
Jeffrey D. Carpenter

1

AUTOMATIC MEASUREMENT OF AUDIO PRESENCE AND LEVEL BY DIRECT PROCESSING OF AN MPEG DATA STREAM

BACKGROUND OF THE INVENTION

5 1. Field of the Invention

The present invention relates to the automatic measurement of audio presence and level by direct processing of an MPEG data stream.

2. Description of the Related Art

10 Digital television, such as that provided by DIRECTV®, the assignee of the present invention, is typically transmitted as a digital data stream encoded using the MPEG (Motion Pictures Experts Group) standard promulgated by the ISO (International Standards Organization). MPEG provides an efficient way to represent video and audio in the form of a compressed bit stream. The MPEG-1 standard is described in a document entitled "Coding

15 of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 MBit/s," ISO/IEC 11172 (1993), which is incorporated by reference herein.

DIRECTV® broadcasts hundreds of channels to its subscribers encoded into different MPEG data streams. However, problems can arise in using these different MPEG data streams, due to the fact that it is difficult to monitor the audio levels of all of the different

20 channels. Thus, the different MPEG data streams may appear to be either too loud or too soft, as compared to other channels, or there may be a loss of audio that is not noticed for some time.

In the prior art, special purpose devices would be used to measure audio levels. However, these special purpose devices require a separate satellite receiver to tune and

25 decode the audio. In addition, these devices generally are not easily integrated into a system architecture in order to control and report and alarm on the measurements.

Consequently, there is a need to monitor the audio levels of an MPEG data stream. Moreover, there is need for the ability to monitor audio levels of MPEG data streams without decompressing the audio data within the MPEG data streams.

30

SUMMARY OF THE INVENTION

The present invention discloses a method, apparatus and article of manufacture for automatic measurements of audio presence and level in an audio signal by direct processing of an MPEG data stream representing the audio signal, without reconstructing the audio signal. Sub-band data is extracted from the data stream, and the extracted sub-band data is dequantized and denormalized. An audio level for the dequantized and denormalized sub-band data is measured without reconstructing the audio signal. Channel characteristics are used in measuring the audio level of the sub-band data, wherein the channel characteristics are used to weight the measured levels. The measured levels are compared against at least one threshold to determine whether an alarm should be triggered.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagram illustrating an overview of a video distribution system according to a preferred embodiment of the present invention;

FIG. 2 is a block diagram that illustrates the structure of an MPEG audio data stream;

FIG. 3 is a diagram illustrating an audio presence and level monitoring system according the preferred embodiment of the present invention;

FIG. 4 is a diagram illustrating the loudness calculation performed by the software program of the audio presence and level monitoring system according the preferred embodiment of the present invention; and

FIG. 5 is a graph showing a plot of audio levels as compared to time performed by an audio presence and level function in the preferred embodiment of the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

In the following description, reference is made to the accompanying drawings that form a part hereof, and which show, by way of illustration, several embodiments of the present invention. It is understood that other embodiments may be utilized and structural changes may be made without departing from the scope of the present invention.

3

Overview

The present invention provides automatic measurements of audio presence and level in an audio signal by direct processing of an MPEG data stream representing the audio signal, without reconstructing the audio signal. Consequently, if the audio level in an MPEG data stream is too high or too low, the audio level can be detected and adjusted as desired in order to maintain uniform listening levels.

A preferred embodiment of the present invention comprises an audio presence and level monitoring system that uses a satellite receiver card connected to a computer system. The satellite receiver card has the capability to receive and process a plurality of MPEG data streams concurrently and make them available to a software program executed by the computer system. The software program calculates the perceived loudness directly from the MPEG data streams without reconstruction of the audio signal.

The present invention can be employed to continuously monitor the audio presence and levels within the MPEG data streams for a video distribution system, such as a satellite broadcast system. Most video distribution systems include many active audio streams and it is important to subscribers that these audio streams are set to the proper level, so as not make channel change objectionable. The audio presence and level monitoring system accomplishes this task with a minimal amount of expense. Moreover, since the system comprises satellite receiver cards integrated into the host computer, the control of the system and the reporting of results are simple to implement.

The present invention improves the quality of services provided to the subscribers of the video distribution system, as well as lowering the cost of providing these services. In particular, the present invention permits audio problems to be recognized and handled more quickly. In the prior art, monitoring of audio presence and levels is a labor intensive activity, since a person assigned to perform this task manually can only listen to one audio stream at a time. For example, in a large multi-channel system, there have been instances where a secondary audio channel has been inoperative for more than a day.

<u>Video Distribution System</u>

FIG. 1 is a diagram illustrating an overview of a video distribution system 100 according to a preferred embodiment of the present invention. The video distribution system 100 comprises a control center 102 in communication with an uplink center 104 via a ground

5   link 106 and with subscriber receiving stations 108 via a link 110. The control center 102 provides program material to the uplink center 104, coordinates with the subscriber receiving stations 108 to offer pay-per-view (PPV) program services, including billing and associated decryption of video programs.

The uplink center 104 receives the program material from the control center 102 and,

10   using an uplink antenna 112 and transmitter 114, transmits the program material to one or more satellites 116, each of which may include one or more transponders 118. The satellites 116 receive and process this information, and transmit the program material to subscriber receiving stations 108 via downlink 120 using transmitter 118. Subscriber receiving stations 108 receive this information using via an antenna 122 of the subscriber receiving stations

15   108.

While the invention disclosed herein will be described with reference to a satellite based video distribution system 100, the present invention may also be practiced with terrestrial-based video distribution system, whether by antenna, cable, or other means. Further, the different functions collectively allocated among the control center 102 and the

20   uplink center 104 as described above can be reallocated as desired without departing from the intended scope of the present invention.

Although the foregoing has been described with respect to an embodiment in which the program material delivered to the subscriber is video (and audio) program material such as a movie, the foregoing method can be used to deliver program material comprising purely

25   audio program material as well. In both instances, the audio program material is encoded as an MPEG audio data stream.

MPEG Audio Data Stream

FIG. 2 is a block diagram that illustrates the structure of an MPEG audio data stream 200. Layers I, II and III within the MPEG audio data stream 200 are shown as separate frames 202, 204 and 206.

Each frame 202, 204 and 206 includes a Header 208, which is followed by an optional cyclic redundancy check (CRC) 210 that is 16 bits in length. The Header 208 is 32 bits and includes the following information:

- Sync Word – 12 bits (all 1s)
- System Word – 20 bits
  - Version id – 1 bit
  - Layer – 2 bits
  - Error Protection – 1 bit
  - Bit Rate Index – 4 bits
  - Sampling Frequency Rate Index – 2 bits
  - Padding – 1 bit
  - Private– 1 bit
  - Mode – 2 bits
  - Mode Extension – 2 bits
  - Copyright – 1 bit
  - Original or copy – 1 bit
  - Emphasis – 2 bits

The CRC 210, if present, is used for detecting errors.

In the frame 202 of Layer 1, the CRC 210 is followed by a Bit Allocation 212 (128-256 bits in length), Scale Factors 214 (0-384 bits in length), Samples 216 (384 bits in length), and Ancillary Data 218. In the frame 204 of Layer 2, the CRC 210 is followed by a Bit Allocation 212 (26-188 bits in length), Scale Factor Selection Information (SCFSI) 220 (0-60 bits in length), Scale Factors 214 (0-1080 bits in length), Samples 216 (1152 bits in length), and Ancillary Data 218. In the frame 206 of Layer 3, the CRC 210 is followed by Side Information 222 (136-256 bits in length) and a Bit Reservoir 224.

In both Layers I and II, the time-frequency mapping of the audio signal uses a polyphase filter bank with 32 sub-bands, wherein the sub-bands are equally spaced in frequency. The Layer 1 psychoacoustic model uses a 512-point Fast Fourier Transform (FFT) to obtain detailed spectral information about the audio signal, while the Layer 2 psychoacoustic model, which is similar to the Layer 1 psychoacoustic model, uses a 1024-point FFT for greater frequency resolution of the audio signal. Both the Layer 1 and II quantizers examine the data in each sub-band to determine the Bit Allocation 212 and Scale Factor 214 for each sub-band, and then linearly quantize the data in each sub-band according to the Bit Allocation 212 and Scale Factor 214 for that sub-band.

The Bit Allocation 212 determines the number of bits per sample for Layer 1, or the number of quantization levels for Layer 2. Specifically, the Bit Allocation 212 specifies the number of bits assigned for quantization of each sub-band. These assignments are made adaptively, according to the information content of the audio signal, so the Bit Allocation 212 varies in each frame 202, 204. The Samples 216 can be coded with zero bits (i.e., no data are present), or with two to fifteen bits per sample.

The Scale Factors 214 are coded to indicate sixty-three possible values that are coded as six-bit index patterns from "000000" (0), which designates the maximum scale factor, to "111110" (62), which designates the minimum scale factor. Each sub-band in the Samples 216 has an associated Scale Factor 214 that defines the level at which each sub-band is recombined during decoding.

The Samples 216 themselves comprise the linearly quanitized data, e.g., samples, for each of thirty-two sub-bands. A Layer 1 frame 202 comprises twelve samples per sub-band. A Layer 2 frame 204 comprises thirty-six samples per sub-band.

In Layer 2 204, the Samples 216 in each frame are divided into three parts, wherein each part comprises twelve samples per sub-band. For each sub-band, the SCFSI 220 indicates whether the three parts have separate Scale Factors 214, or all three parts have the same Scale Factor 214, or two parts (the first two or the last two) have one Scale Factor 214 and the other part has another Scale Factor 214.

Monitoring System

FIG. 3 is a diagram illustrating a audio presence and level monitoring system 300 according the preferred embodiment of the present invention. The audio presence and level monitoring system 300 may comprise one or more of the subscriber receiver stations 108

5 described in FIG. 1, although the audio presence and level monitoring system 300 may also comprise a component of the control center 102 or uplink center 104 described in FIG. 1. Indeed, the audio presence and level monitoring system 300 may be located wherever it is convenient for monitoring MPEG audio data streams.

The audio presence and level monitoring system 300 includes one or more host

10 computers 302, each of which includes at least one satellite receiver card 304 and software program 306 executed by the host computer 300, or alternatively, by the satellite receiver card 304. The satellite receiver card 304 is coupled to an L-band distribution device 308, which includes a satellite dish 310, in order to receive one or more MPEG audio data streams. A system configuration and management system 312 is used to configure and

15 manage the host computers 302 and satellite receiver cards 304, while an error monitoring system 314 is notified if any errors are detected in the MPEG audio data streams.

Preferably, the satellite receiver card 304 has the capability to tune a plurality of MPEG audio data streams at once. The MPEG audio data streams are then transferred directly to memory in the host computer 302.

20 The software program 306 comprises an MPEG audio parser that accesses the MPEG audio data stream from the memory and rebuilds the data therein into Layer 2 frames 204, in order to access a set of sub-bands in the Samples 216. However, instead of reconstructing the audio signal, which would complete the normal decoding process for the MPEG audio data stream 200, the sub-band data in the Samples 216 are processed in an audio presence and

25 level function performed by the software program 306, since the sub-band data is already represented in a fashion that is easily scalable for the human ear sensitivity.

The audio presence and level function performed by the software program 306 typically involves measuring the power of the sub-band data in the Samples 216, wherein the power is measured as a square root of a sum of squares of the sub-band data. The software

30 program 306 then averages and thresholds the measured power over time to calculate the

8

level. However, the details of the calculation may vary by application.

For audio presence detection, the signal power will go to zero (or below a low threshold). The channel characteristics can be used to determine the length of time the signal power is at zero before considering the audio signal as being lost. Moreover, thresholds can
5   be set to generate an alarm based on loss of the audio signal, or when the average level of the audio signal is too high or too low.

As most of the processing is in the software program 306, there is no need for specialized hardware and wiring for each channel. In one embodiment, the system 300 is capable of measuring all the MPEG audio data streams on a transponder with one satellite
10   receiver card 304, which entails processing 20-30 channels in parallel per satellite receiver card 304.

In one embodiment, a single satellite receiver card 304 can be dedicated to each transponder for full time monitoring. On the other hand, if each of the MPEG audio data streams need only to be sampled occasionally, then in an alternative embodiment, a single
15   satellite receiver card 304 can be re-tuned to different transponders, to sample different channels, and monitor an entire video broadcast system by cycling through a full set (e.g., 40-60) of transponders.

Audio Presence and Level Function

20

1.   A method of automatic measurement of audio presence and level by direct processing of a data stream representing an audio signal, comprising:

FIG. 4 is a diagram illustrating the audio presence and level function performed by the software program 306 of the audio presence and level monitoring system 300 according the
25   preferred embodiment of the present invention. Specifically, the blocks of the diagram represent a method of automatic measurement of audio presence and level by direct processing of an MPEG data stream representing an audio signal.

Initially, the sub-band data 400 is extracted from the MPEG data stream. The values of the sub-band data 400 represents the strength of the audio signal in a frequency band
30   covered by the sub-band data 400 at that point in time.

9

A hearing curve scaling function 402 performs the steps of dequantizing and denormalizing the extracted sub-band data, wherein the sub-band data 400 is dequantized according to the Bit Allocation 212, and the sub-band data 400 is denormalized using the Scale Factors 214. However, there is no need to further reconstruct the audio signal, which would complete the normal decoding process for the MPEG audio data stream 200. The sub-band data 400 is already in the frequency domain, so it is easily scalable to compensate for human ear sensitivity.

The hearing curve scaling function 402 also performs the step of using a psychoacoustic model 404 in determining a perceived level of the measured audio signal according to human ear sensitivity. Human ear sensitivity is frequency dependent and is more sensitive to mid-range frequencies (1-3 kHz) than to low or high frequency signals

A loudness calculation 406 performs the steps of measuring an audio level for the dequantized and denormalized sub-band data 400 without reconstructing the audio signal using one or more channel characteristics 408, averaging the measured audio levels over time, and comparing the averaged audio levels against at least one application-specific threshold to determine whether the threshold is exceeded. The loudness calculation 406 typically involves measuring the signal power of the audio signal, as represented by the sub-band data 400, to determine the audio presence and level. However, the details of the calculation may vary by application.

The channel characteristics 408 are used to weight an instantaneous level or an overall level for a channel. For example, commercial advertising material typically has a different perceived level from the nominal program material, and it might be useful to exclude the instantaneous level for the commercial advertising material from the overall perceived level for a channel, because the commercial advertising material is normally a small percentage of total audio content. The exclusion can be accomplished using the channel characteristics 408, which may comprise a schedule for commercial breaks, or a label for the channel, or a label for the MPEG audio data stream, or some other indicator.

When a threshold is exceeded, an alarm 410 is triggered and passed onto the error monitoring system 312. Based on the alarm 410, one or more actions can be taken, i.e., adjusting audio levels, or tracing out the lost audio signal, or some other action. For

10

example, a Simple Network Management Protocol (SNMP) agent may be used to report alarms, levels or other information as they occur.

FIG. 5 is a graph showing a plot 500 of audio levels as compared to time performed by the audio presence and level function of the software program 306 according to the preferred embodiment of the present invention. In the plot 500, the audio levels as compared to time are compared to a high threshold 502, low threshold 504 and presence threshold 506.

The thresholds 502, 504, and 506 can be determined by experimentation and tuned to yield the desired results. The audio presence and level function does not have to be perfect, but instead, only needs to be monotonic and reasonably linear.

In addition, the thresholds 502, 504, and 506 may vary from channel to channel depending on the channel characteristics 408 of a channel. Moreover, the channel characteristics 408 may vary over the course of the day and so the thresholds 502, 504, and 506 may also be varied by time of day, if that proves to be appropriate.

<u>Conclusion</u>

The foregoing description of the preferred embodiment of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching.

For example, while the foregoing disclosure presents an embodiment of the present invention as it is applied to a satellite transmission system, the present invention can be applied to any application that uses MPEG audio. Moreover, although the present invention is described in terms of MPEG audio, it could also be applied to other compression schemes, such as Dolby® AC-3. Finally, although specific hardware, software and logic is described herein, those skilled in the art will recognize that other hardware, software or logic may accomplish the same result, without departing from the scope of the present invention.

It is intended that the scope of the invention be limited not by this detailed description, but rather by the claims appended hereto. The above specification, examples and data provide a complete description of the manufacture and use of the composition of the

invention. Since many embodiments of the invention can be made without departing from the spirit and scope of the invention, the invention resides in the claims hereinafter appended.